Concentration Results For Heavy-tailed Random Distributions

INFORMS 2020

Milad Bakhshizadeh, Arian Maleki, Victor De la Pena

Columbia University

November 8, 2020

Inverse Problem

- Find $\boldsymbol{x}$ s.t. $y_i = A(\boldsymbol{x}, \boldsymbol{a}_i), \ i = 1, 2, ..., n$
- $f(\boldsymbol{c}) = \sum_i \left\| A(\boldsymbol{c}, \boldsymbol{a}_i) - y_i \right\|^2, \quad \hat{\boldsymbol{x}} \triangleq \underset{\boldsymbol{c} \in D \subset \mathbb{R}^n}{\operatorname{argmin}} f(\boldsymbol{c})$
  - $\boldsymbol{x}_0, \quad \boldsymbol{x}_+ = \boldsymbol{x} - \mu \nabla f(\boldsymbol{x})$
  - $\boldsymbol{x}_+ \simeq \boldsymbol{x} - \mu \mathbb{E}\left(\nabla f(\boldsymbol{x})\right)$
  - $\left\| \nabla f - \mathbb{E}\left(\nabla f\right) \right\|, \ \nabla f(\boldsymbol{x}) = \sum_i \nabla \left\| (A(\boldsymbol{x}, \boldsymbol{a}_i)) - y_i \right\|^2 = \sum_i \boldsymbol{z}_i$
- $\boldsymbol{z}_i \overset{iid}{\sim} \boldsymbol{z}, \quad \mathbb{P}\left(\left\| \frac{1}{n} \sum \boldsymbol{z}_i - \mathbb{E}\left(\boldsymbol{z}\right) \right\| > \epsilon \right)$
- Convergence rate, Sample Complexity

Inverse Problem

- Find $\boldsymbol{x}$ s.t. $y_i = A(\boldsymbol{x}, \boldsymbol{a}_i), \; i = 1, 2, \ldots, n$
- $f(\boldsymbol{c}) = \sum_i \left\| A(\boldsymbol{c}, \boldsymbol{a}_i) - y_i \right\|^2, \quad \hat{\boldsymbol{x}} \triangleq \underset{\boldsymbol{c} \in D \subset \mathbb{R}^n}{\operatorname{argmin}} f(\boldsymbol{c})$
  - $\boldsymbol{x}_0, \quad \boldsymbol{x}_+ = \boldsymbol{x} - \mu \nabla f(\boldsymbol{x})$
  - $\boldsymbol{x}_+ \simeq \boldsymbol{x} - \mu \mathbb{E}\left(\nabla f(\boldsymbol{x})\right)$
  - $\left\| \nabla f - \mathbb{E}\left(\nabla f\right) \right\|, \; \nabla f(\boldsymbol{x}) = \sum_i \nabla \left\| \left(A(\boldsymbol{x}, \boldsymbol{a}_i)\right) - y_i \right\|^2 = \sum_i \boldsymbol{z}_i$
- $\boldsymbol{z}_i \overset{iid}{\sim} \boldsymbol{z}, \quad \mathbb{P}\left(\left\| \frac{1}{n} \sum \boldsymbol{z}_i - \mathbb{E}\left(\boldsymbol{z}\right) \right\| > \epsilon \right)$
- Convergence rate, Sample Complexity

Inverse Problem

- Find $\boldsymbol{x}$ s.t. $y_i = A(\boldsymbol{x}, \boldsymbol{a}_i),\ i = 1, 2, ..., n$
- $f(\boldsymbol{c}) = \sum_i \left\| A(\boldsymbol{c}, \boldsymbol{a}_i) - y_i \right\|^2, \quad \hat{\boldsymbol{x}} \triangleq \operatorname*{argmin}_{\boldsymbol{c} \in D \subset \mathbb{R}^n} f(\boldsymbol{c})$
  - $\boldsymbol{x}_0, \quad \boldsymbol{x}_+ = \boldsymbol{x} - \mu \nabla f(\boldsymbol{x})$
  - $\boldsymbol{x}_+ \simeq \boldsymbol{x} - \mu \mathbb{E}\left(\nabla f(\boldsymbol{x})\right)$
  - $\left\| \nabla f - \mathbb{E}(\nabla f) \right\|,\ \nabla f(\boldsymbol{x}) = \sum_i \nabla \left\| (A(\boldsymbol{x}, \boldsymbol{a}_i)) - y_i \right\|^2 = \sum_i \boldsymbol{z}_i$
- $\boldsymbol{z}_i \overset{iid}{\sim} \boldsymbol{z}, \quad \mathbb{P}\left(\left\| \frac{1}{n} \sum \boldsymbol{z}_i - \mathbb{E}(\boldsymbol{z}) \right\| > \epsilon \right)$
- Convergence rate, Sample Complexity

Inverse Problem

- Find $\boldsymbol{x}$ s.t. $y_i = A(\boldsymbol{x}, \boldsymbol{a}_i), \; i = 1, 2, ..., n$
- $f(\boldsymbol{c}) = \sum_i \left\| A(\boldsymbol{c}, \boldsymbol{a}_i) - y_i \right\|^2, \quad \hat{\boldsymbol{x}} \triangleq \underset{\boldsymbol{c} \in D \subset \mathbb{R}^n}{\operatorname{argmin}} f(\boldsymbol{c})$
  - $\boldsymbol{x}_0, \quad \boldsymbol{x}_+ = \boldsymbol{x} - \mu \nabla f(\boldsymbol{x})$
  - $\boldsymbol{x}_+ \simeq \boldsymbol{x} - \mu \mathbb{E}\left(\nabla f(\boldsymbol{x})\right)$
  - $\left\| \nabla f - \mathbb{E}(\nabla f) \right\|, \; \nabla f(\boldsymbol{x}) = \sum_i \nabla \left\| (A(\boldsymbol{x}, \boldsymbol{a}_i)) - y_i \right\|^2 = \sum_i \boldsymbol{z}_i$
- $\boldsymbol{z}_i \overset{iid}{\sim} \boldsymbol{z}, \quad \mathbb{P}\left(\left\| \frac{1}{n} \sum \boldsymbol{z}_i - \mathbb{E}(\boldsymbol{z}) \right\| > \epsilon \right)$
- Convergence rate, Sample Complexity

Inverse Problem

- Find $\boldsymbol{x}$ s.t. $y_i = A(\boldsymbol{x}, \boldsymbol{a}_i),\ i = 1, 2, ..., n$
- $f(\boldsymbol{c}) = \sum_i \left\| A(\boldsymbol{c}, \boldsymbol{a}_i) - y_i \right\|^2, \quad \hat{\boldsymbol{x}} \triangleq \underset{\boldsymbol{c} \in D \subset \mathbb{R}^n}{\text{argmin}}\ f(\boldsymbol{c})$
  - $\boldsymbol{x}_0, \quad \boldsymbol{x}_+ = \boldsymbol{x} - \mu \nabla f(\boldsymbol{x})$
  - $\boldsymbol{x}_+ \simeq \boldsymbol{x} - \mu \mathbb{E}\left(\nabla f(\boldsymbol{x})\right)$
  - $\left\| \nabla f - \mathbb{E}\left(\nabla f\right) \right\|,\ \nabla f(\boldsymbol{x}) = \sum_i \nabla \left\| \left(A(\boldsymbol{x}, \boldsymbol{a}_i)\right) - y_i \right\|^2 = \sum_i \boldsymbol{z}_i$
- $\boldsymbol{z}_i \overset{iid}{\sim} \boldsymbol{z}, \quad \mathbb{P}\left(\left\| \frac{1}{n} \sum \boldsymbol{z}_i - \mathbb{E}(\boldsymbol{z}) \right\| > \epsilon \right)$
- Convergence rate, Sample Complexity

Inverse Problem

- Find $\boldsymbol{x}$ s.t. $y_i = A(\boldsymbol{x}, \boldsymbol{a}_i),\ i = 1, 2, ..., n$
- $f(\boldsymbol{c}) = \sum_i \left\| A(\boldsymbol{c}, \boldsymbol{a}_i) - y_i \right\|^2, \quad \hat{\boldsymbol{x}} \triangleq \underset{\boldsymbol{c} \in D \subset \mathbb{R}^n}{\operatorname{argmin}} f(\boldsymbol{c})$
  - $\boldsymbol{x}_0, \quad \boldsymbol{x}_+ = \boldsymbol{x} - \mu \nabla f(\boldsymbol{x})$
  - $\boldsymbol{x}_+ \simeq \boldsymbol{x} - \mu \mathbb{E}\left(\nabla f(\boldsymbol{x})\right)$
  - $\left\| \nabla f - \mathbb{E}(\nabla f) \right\|, \ \nabla f(\boldsymbol{x}) = \sum_i \nabla \left\| \left(A(\boldsymbol{x}, \boldsymbol{a}_i)\right) - y_i \right\|^2 = \sum_i \boldsymbol{z}_i$
- $\boldsymbol{z}_i \overset{iid}{\sim} \boldsymbol{z}, \quad \mathbb{P}\left(\left\| \frac{1}{n} \sum \boldsymbol{z}_i - \mathbb{E}(\boldsymbol{z}) \right\| > \epsilon\right)$
- Convergence rate, Sample Complexity

Inverse Problem

- Find $\boldsymbol{x}$ s.t. $y_i = A(\boldsymbol{x}, \boldsymbol{a}_i), \ i = 1, 2, ..., n$
- $f(\boldsymbol{c}) = \sum_i \left\| A(\boldsymbol{c}, \boldsymbol{a}_i) - y_i \right\|^2, \quad \hat{\boldsymbol{x}} \triangleq \underset{\boldsymbol{c} \in D \subset \mathbb{R}^n}{\operatorname{argmin}} f(\boldsymbol{c})$
  - $\boldsymbol{x}_0, \quad \boldsymbol{x}_+ = \boldsymbol{x} - \mu \nabla f(\boldsymbol{x})$
  - $\boldsymbol{x}_+ \simeq \boldsymbol{x} - \mu \mathbb{E}\left(\nabla f(\boldsymbol{x})\right)$
  - $\left\| \nabla f - \mathbb{E}(\nabla f) \right\|, \ \nabla f(\boldsymbol{x}) = \sum_i \nabla \left\| \left(A(\boldsymbol{x}, \boldsymbol{a}_i)\right) - y_i \right\|^2 = \sum_i \boldsymbol{z}_i$
- $\boldsymbol{z}_i \overset{iid}{\sim} \boldsymbol{z}, \quad \mathbb{P}\left(\left\| \frac{1}{n} \sum \boldsymbol{z}_i - \mathbb{E}(\boldsymbol{z}) \right\| > \epsilon \right)$
- Convergence rate, Sample Complexity

Concentration of Measure

- $z_i \overset{iid}{\sim} z \in \mathbb{R}, \quad \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right) \leq U_z(\epsilon, n)$
  - $\lim_{n\to\infty} U_z(\epsilon, n) = 0$ (LLN)
  - $\epsilon = \epsilon(n) \to 0$
- Sharpness
  - $L_z(\epsilon, n) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right)$
  - $\lim_{n\to\infty} \frac{U_z(\epsilon,n)}{L_z(\epsilon,n)} = 1$

Concentration of Measure

- $z_i \overset{iid}{\sim} z \in \mathbb{R}, \quad \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right) \leq U_z(\epsilon, n)$
  - $\lim_{n \to \infty} U_z(\epsilon, n) = 0$ (LLN)
  - $\epsilon = \epsilon(n) \to 0$
- Sharpness
  - $L_z(\epsilon, n) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right)$
  - $\lim_{n \to \infty} \frac{U_z(\epsilon, n)}{L_z(\epsilon, n)} = 1$

Concentration of Measure

- $z_i \overset{iid}{\sim} z \in \mathbb{R}, \quad \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right) \leq U_z(\epsilon, n)$
  - $\lim_{n\to\infty} U_z(\epsilon, n) = 0$ (LLN)
  - $\epsilon = \epsilon(n) \to 0$
- Sharpness
  - $L_z(\epsilon, n) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right)$
  - $\lim_{n\to\infty} \frac{U_z(\epsilon,n)}{L_z(\epsilon,n)} = 1$

Concentration of Measure

- $z_i \overset{iid}{\sim} z \in \mathbb{R}, \quad \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right) \leq U_z(\epsilon, n)$
  - $\lim_{n\to\infty} U_z(\epsilon, n) = 0$ (LLN)
  - $\epsilon = \epsilon(n) \to 0$
- Sharpness
  - $L_z(\epsilon, n) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right)$
  - $\lim_{n\to\infty} \frac{U_z(\epsilon,n)}{L_z(\epsilon,n)} = 1$

Concentration of Measure

- $z_i \overset{iid}{\sim} z \in \mathbb{R}, \quad \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right) \leq U_z(\epsilon, n)$
  - $\lim_{n\to\infty} U_z(\epsilon, n) = 0$ (LLN)
  - $\epsilon = \epsilon(n) \to 0$
- Sharpness
  - $L_z(\epsilon, n) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right)$
  - $\lim_{n\to\infty} \frac{U_z(\epsilon,n)}{L_z(\epsilon,n)} = 1$

Concentration of Measure

- $z_i \overset{iid}{\sim} z \in \mathbb{R}, \quad \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right) \leq U_z(\epsilon, n)$
  - $\lim_{n\to\infty} U_z(\epsilon, n) = 0$ (LLN)
  - $\epsilon = \epsilon(n) \to 0$
- Sharpness
  - $L_z(\epsilon, n) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum z_i - \mathbb{E}(z)\right| > \epsilon\right)$
  - $\lim_{n\to\infty} \frac{U_z(\epsilon,n)}{L_z(\epsilon,n)} = 1$

Higher Dimensions

- Reduced to $\mathbb{R}$ by Projection
  - $\|\boldsymbol{z}\| = \max_{\|\boldsymbol{v}\|=1} \langle \boldsymbol{z}, \boldsymbol{v} \rangle$
- ϵ-net
  - $\mathcal{B}(\mathbf{0}, 1) \subseteq \cup_{i=1}^{k} \mathcal{B}(\boldsymbol{v}_i, \epsilon)$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}(z) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = ke^{-c_z\epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}(z) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k \max\left(e^{-c_z^2\epsilon^2 n}, e^{-c_z\epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = k\mathrm{e}^{-c_z \epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k \max\left(\mathrm{e}^{-c_z^2 \epsilon^2 n}, \mathrm{e}^{-c_z \epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = k\mathrm{e}^{-c_z\epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k \max\left(\mathrm{e}^{-c_z^2\epsilon^2 n}, \mathrm{e}^{-c_z\epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = k\mathrm{e}^{-c_z\epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k\max\left(\mathrm{e}^{-c_z^2\epsilon^2 n}, \mathrm{e}^{-c_z\epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = k\mathrm{e}^{-c_z \epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k \max\left(\mathrm{e}^{-c_z^2 \epsilon^2 n}, \mathrm{e}^{-c_z \epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = k\mathrm{e}^{-c_z\epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k \max\left(\mathrm{e}^{-c_z^2\epsilon^2 n}, \mathrm{e}^{-c_z\epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}(z) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = k\mathrm{e}^{-c_z\epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}(z) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k \max\left(\mathrm{e}^{-c_z^2\epsilon^2 n}, \mathrm{e}^{-c_z\epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = k\mathrm{e}^{-c_z\epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k\max\left(\mathrm{e}^{-c_z^2\epsilon^2 n}, \mathrm{e}^{-c_z\epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

SubGaussians and subexponentials

- Role of the tail
- SubGaussian: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct^2$
  - Normal, Bernoulli, Bounded
  - General Hoeffding's Ineq. $U_z(\epsilon, n) = k\mathrm{e}^{-c_z\epsilon^2 n}$
  - $\epsilon \gg n^{-\frac{1}{2}}$
- Subexponential: $\log \mathbb{P}\left(z - \mathbb{E}\left(z\right) > t\right) \sim -ct$
  - χ^2, Exponential
  - Bernsteins Ineq. $U_z(\epsilon, n) = k \max\left(\mathrm{e}^{-c_z^2\epsilon^2 n}, \mathrm{e}^{-c_z\epsilon n}\right)$
  - SubGaussian Region: $\epsilon \leq c_z$

Proof technique (classical concentration inequalities)

- $\mathbb{E}(z) = 0$
- Moment Generating Function
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$ (Markov's)
- Optimize λ

Proof technique (classical concentration inequalities)

- $\mathbb{E}(z) = 0$
- Moment Generating Function
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$ (Markov's)
- Optimize λ

Proof technique (classical concentration inequalities)

- $\mathbb{E}(z) = 0$
- Moment Generating Function
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$ (Markov's)
- Optimize λ

Proof technique (classical concentration inequalities)

- $\mathbb{E}(z) = 0$
- Moment Generating Function
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$ (Markov's)
- Optimize λ

Heavier tails

- Rate function: $I(z) \triangleq -\log \mathbb{P}(z > t)$
  - (SG): $I(z) \sim z^2$, (SE): $I(z) \sim z$, Heavy tails: $I(z) \ll z$
  - Multiplication
  - $\mathcal{N}(0,1)^p, \ p > 2$
- MGF
  - $I(z) \sim z^2, z$
  - $I(z) \ll z \implies \mathrm{MGF}(\lambda) = \infty \ \forall \lambda$
  - Classical proof does not work

Heavier tails

- Rate function: $I(z) \triangleq -\log \mathbb{P}(z > t)$
  - (SG): $I(z) \sim z^2$, (SE): $I(z) \sim z$, Heavy tails: $I(z) \ll z$
  - Multiplication
  - $\mathcal{N}(0,1)^p, \; p > 2$
- MGF
  - $I(z) \sim z^2, z$
  - $I(z) \ll z \implies \mathrm{MGF}(\lambda) = \infty \; \forall \lambda$
  - Classical proof does not work

Heavier tails

- Rate function: $I(z) \triangleq -\log \mathbb{P}(z > t)$
  - (SG): $I(z) \sim z^2$, (SE): $I(z) \sim z$, Heavy tails: $I(z) \ll z$
  - Multiplication
  - $\mathcal{N}(0,1)^p, \ p > 2$
- MGF
  - $I(z) \sim z^2, z$
  - $I(z) \ll z \implies \mathrm{MGF}(\lambda) = \infty \ \forall \lambda$
  - Classical proof does not work

Milad Bakhshizadeh, Arian Maleki, Victor De la Pena
Columbia University

Heavier tails

- Rate function: $I(z) \triangleq -\log \mathbb{P}\left(z > t\right)$
  - (SG): $I(z) \sim z^2$, (SE): $I(z) \sim z$, Heavy tails: $I(z) \ll z$
  - Multiplication
  - $\mathcal{N}(0,1)^p, \; p > 2$
- MGF
  - $I(z) \sim z^2, z$
  - $I(z) \ll z \implies \mathrm{MGF}(\lambda) = \infty \; \forall \lambda$
  - Classical proof does not work

Heavier tails

- Rate function: $I(z) \triangleq -\log \mathbb{P}(z > t)$
  - (SG): $I(z) \sim z^2$, (SE): $I(z) \sim z$, Heavy tails: $I(z) \ll z$
  - Multiplication
  - $\mathcal{N}(0,1)^p, \; p > 2$
- MGF
  - $I(z) \sim z^2, z$
  - $I(z) \ll z \implies \mathrm{MGF}(\lambda) = \infty \; \forall \lambda$
  - Classical proof does not work

Heavier tails

- Rate function: $I(z) \triangleq -\log \mathbb{P}(z > t)$
  - (SG): $I(z) \sim z^2$, (SE): $I(z) \sim z$, Heavy tails: $I(z) \ll z$
  - Multiplication
  - $\mathcal{N}(0,1)^p, \ p > 2$
- MGF
  - $I(z) \sim z^2, z$
  - $I(z) \ll z \implies \mathrm{MGF}(\lambda) = \infty \ \forall \lambda$
  - Classical proof does not work

Heavier tails

- Rate function: $I(z) \triangleq -\log \mathbb{P}\,(z > t)$
  - (SG): $I(z) \sim z^2$, (SE): $I(z) \sim z$, Heavy tails: $I(z) \ll z$
  - Multiplication
  - $\mathcal{N}(0,1)^p, \; p > 2$
- MGF
  - $I(z) \sim z^2, z$
  - $I(z) \ll z \implies \mathrm{MGF}(\lambda) = \infty \; \forall \lambda$
  - Classical proof does not work

Heavier tails

- Rate function: $I(z) \triangleq -\log \mathbb{P}(z > t)$
  - (SG): $I(z) \sim z^2$, (SE): $I(z) \sim z$, Heavy tails: $I(z) \ll z$
  - Multiplication
  - $\mathcal{N}(0,1)^p, \; p > 2$
- MGF
  - $I(z) \sim z^2, z$
  - $I(z) \ll z \implies \mathrm{MGF}(\lambda) = \infty \; \forall \lambda$
  - Classical proof does not work

Related Works

- Probability and Finance
  - Nagaev, Sergey V. "Large deviations of sums of independent random variables." The Annals of Probability (1979): 745-789.
  - Klass, M., & Nowicki, K. (2007). Uniformly accurate quantile bounds via the truncated moment generating function: the symmetric case. Electronic Journal of Probability, 12, 1276-1298.
  - Too general, Free parameters, Several Inequalities
- Statistics (Recent)
  - Kuchibhotla, A. K., & Chakrabortty, A. (2018). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression.
  - Sambale, H. (2020). Some notes on concentration for α-subexponential random variables.
  - Orlicz norm, SubWeibull

Related Works

- Probability and Finance
  - Nagaev, Sergey V. "Large deviations of sums of independent random variables." The Annals of Probability (1979): 745-789.
  - Klass, M., & Nowicki, K. (2007). Uniformly accurate quantile bounds via the truncated moment generating function: the symmetric case. Electronic Journal of Probability, 12, 1276-1298.
  - Too general, Free parameters, Several Inequalities
- Statistics (Recent)
  - Kuchibhotla, A. K., & Chakrabortty, A. (2018). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression.
  - Sambale, H. (2020). Some notes on concentration for α-subexponential random variables.
  - Orlicz norm, SubWeibull

Related Works

- Probability and Finance
  - Nagaev, Sergey V. "Large deviations of sums of independent random variables." The Annals of Probability (1979): 745-789.
  - Klass, M., & Nowicki, K. (2007). Uniformly accurate quantile bounds via the truncated moment generating function: the symmetric case. Electronic Journal of Probability, 12, 1276-1298.
  - Too general, Free parameters, Several Inequalities
- Statistics (Recent)
  - Kuchibhotla, A. K., & Chakrabortty, A. (2018). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression.
  - Sambale, H. (2020). Some notes on concentration for α-subexponential random variables.
  - Orlicz norm, SubWeibull

Related Works

- Probability and Finance
  - Nagaev, Sergey V. "Large deviations of sums of independent random variables." The Annals of Probability (1979): 745-789.
  - Klass, M., & Nowicki, K. (2007). Uniformly accurate quantile bounds via the truncated moment generating function: the symmetric case. Electronic Journal of Probability, 12, 1276-1298.
  - Too general, Free parameters, Several Inequalities
- Statistics (Recent)
  - Kuchibhotla, A. K., & Chakrabortty, A. (2018). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression.
  - Sambale, H. (2020). Some notes on concentration for α-subexponential random variables.
  - Orlicz norm, SubWeibull

Inequalities for heavy tails

- Rate function: $-\log \mathbb{P}\left(z > t\right)$
- Capture Right Tail: $\mathbb{P}\left(z > t\right) \leq \mathrm{e}^{-I(t)}, \quad \forall t > 0$
- $0 < \beta \leq 1$
- $c = c\left(n, \epsilon\right), \ \frac{1}{2} \leq c \leq 1, \quad c \xrightarrow{n \to \infty} 1$

[1] Bakhshizadeh, M., Maleki, A. and De la Pena, V.H., 2020. Sharp Concentration Results for Heavy-Tailed Distributions.

Inequalities for heavy tails

- Rate function: $-\log \mathbb{P}(z > t)$
- Capture Right Tail: $\mathbb{P}(z > t) \leq \mathrm{e}^{-I(t)}, \quad \forall t > 0$

- $0 < \beta \leq 1$
- $c = c(n, \epsilon), \ \frac{1}{2} \leq c \leq 1, \quad c \xrightarrow{n \to \infty} 1$

[1] Bakhshizadeh, M., Maleki, A. and De la Pena, V.H., 2020. Sharp Concentration Results for Heavy-Tailed Distributions.

Inequalities for heavy tails

- Rate function: $-\log \mathbb{P}(z > t)$
- Capture Right Tail: $\mathbb{P}(z > t) \leq \mathrm{e}^{-I(t)}, \quad \forall t > 0$

Thm (simplified) [1]

$$\mathbb{P}\left(\frac{\sum z_i}{n} > \epsilon\right) \leq \begin{cases} \exp\left(-c\beta I(n\epsilon)\right) + n\exp\left(-I(n\epsilon)\right), & \epsilon \geq \epsilon_0, \\ \exp\left(-\frac{n\epsilon^2}{2c_{n\epsilon_0,\beta}}\right) + n\exp\left(-\frac{n\epsilon_0^2}{\beta c_{n\epsilon_0,\beta}}\right), & 0 \leq \epsilon < \epsilon_0. \end{cases}$$

- $0 < \beta \leq 1$
- $c = c(n,\epsilon), \ \frac{1}{2} \leq c \leq 1, \quad c \xrightarrow{n\to\infty} 1$

[1] Bakhshizadeh, M., Maleki, A. and De la Pena, V.H., 2020. Sharp Concentration Results for Heavy-Tailed Distributions.

Inequalities for heavy tails

- Rate function: $-\log \mathbb{P}(z > t)$
- Capture Right Tail: $\mathbb{P}(z > t) \leq \mathrm{e}^{-I(t)}, \quad \forall t > 0$

Thm (simplified) [1]

$$\mathbb{P}\left(\frac{\sum z_i}{n} > \epsilon\right) \leq \begin{cases} \exp\left(-c\beta I(n\epsilon)\right) + n\exp\left(-I(n\epsilon)\right), & \epsilon \geq \epsilon_0, \\ \exp\left(-\frac{n\epsilon^2}{2c_{n\epsilon_0,\beta}}\right) + n\exp\left(-\frac{n\epsilon_0^2}{\beta c_{n\epsilon_0,\beta}}\right), & 0 \leq \epsilon < \epsilon_0. \end{cases}$$

- $0 < \beta \leq 1$
- $c = c(n,\epsilon), \ \frac{1}{2} \leq c \leq 1, \quad c \xrightarrow{n\to\infty} 1$

[1] Bakhshizadeh, M., Maleki, A. and De la Pena, V.H., 2020. Sharp Concentration Results for Heavy-Tailed Distributions.

Inequalities for heavy tails

- Rate function: $-\log \mathbb{P}(z > t)$
- Capture Right Tail: $\mathbb{P}(z > t) \leq \mathrm{e}^{-I(t)}, \quad \forall t > 0$

Thm (simplified) [1]

$$\mathbb{P}\left(\frac{\sum z_i}{n} > \epsilon\right) \leq \begin{cases} \exp\left(-c\beta I(n\epsilon)\right) + n\exp\left(-I(n\epsilon)\right), & \epsilon \geq \epsilon_0, \\ \exp\left(-\frac{n\epsilon^2}{2c_{n\epsilon_0,\beta}}\right) + n\exp\left(-\frac{n\epsilon_0^2}{\beta c_{n\epsilon_0,\beta}}\right), & 0 \leq \epsilon < \epsilon_0. \end{cases}$$

- $0 < \beta \leq 1$
- $c = c(n, \epsilon), \ \frac{1}{2} \leq c \leq 1, \quad c \xrightarrow{n \to \infty} 1$

[1] Bakhshizadeh, M., Maleki, A. and De la Pena, V.H., 2020. Sharp Concentration Results for Heavy-Tailed Distributions.

Inequalities for heavy tails

- Rate function: $-\log \mathbb{P}(z > t)$
- Capture Right Tail: $\mathbb{P}(z > t) \leq e^{-I(t)}, \quad \forall t > 0$

Thm (simplified) [1]

$$\mathbb{P}\left(\frac{\sum z_i}{n} > \epsilon\right) \leq \begin{cases} \exp\left(-c\beta I(n\epsilon)\right) + n\exp\left(-I(n\epsilon)\right), & \epsilon \geq \epsilon_0, \\ \exp\left(-\frac{n\epsilon^2}{2c_{n\epsilon_0,\beta}}\right) + n\exp\left(-\frac{n\epsilon_0^2}{\beta c_{n\epsilon_0,\beta}}\right), & 0 \leq \epsilon < \epsilon_0. \end{cases}$$

- $0 < \beta \leq 1$
- $c = c(n,\epsilon), \ \frac{1}{2} \leq c \leq 1, \quad c \xrightarrow{n\to\infty} 1$

[1] Bakhshizadeh, M., Maleki, A. and De la Pena, V.H., 2020. Sharp Concentration Results for Heavy-Tailed Distributions.

Applicability

- bound $c_{L,\beta}$
- $c_{L,\beta} = \mathbb{E}\left(\left(z^L\right)^2 \mathbb{I}\left(z^L \leq 0\right) + \left(z^L\right)^2 \exp\left(\frac{\beta I(L)}{L} z^L\right) \mathbb{I}\left(z^L > 0\right)\right)$
  - $z^L = z\mathbb{I}(z \leq L)$
  - $L = n\epsilon \to \infty$
  - $0 < \beta \leq 1$
- $\mathrm{Var}(z) < \infty$

Applicability

- bound $c_{L,\beta}$
- $c_{L,\beta} = \mathbb{E}\left(\left(z^L\right)^2 \mathbb{I}\left(z^L \leq 0\right) + \left(z^L\right)^2 \exp\left(\frac{\beta I(L)}{L} z^L\right) \mathbb{I}\left(z^L > 0\right)\right)$
  - $z^L = z\mathbb{I}(z \leq L)$
  - $L = n\epsilon \to \infty$
  - $0 < \beta \leq 1$
- $\mathrm{Var}(z) < \infty$

Applicability

- bound $c_{L,\beta}$
- $c_{L,\beta} = \mathbb{E}\left(\left(z^L\right)^2 \mathbb{I}\left(z^L \leq 0\right) + \left(z^L\right)^2 \exp\left(\frac{\beta I(L)}{L} z^L\right) \mathbb{I}\left(z^L > 0\right)\right)$
  - $z^L = z\mathbb{I}(z \leq L)$
  - $L = n\epsilon \to \infty$
  - $0 < \beta \leq 1$
- $\mathrm{Var}(z) < \infty$

More technical terms

- $c = c(n, \epsilon, \beta) = 1 - \frac{1}{2} \frac{\beta c_{n\epsilon,\beta}}{\epsilon} \frac{I(n\epsilon)}{n\epsilon}$
- $\epsilon_0 = \epsilon_0(n, \epsilon, \beta) = \sup \left\{ e \geq 0 : \ e \leq \beta c_{n\epsilon,\beta} \frac{I(n\epsilon)}{n\epsilon} \right\}$

Sharpness

- Asymptotic Sharpness: Fix ϵ, $n \to \infty$
  - First Region: $\epsilon \geq \epsilon_0$
    $\mathbb{P}\left(\frac{\sum z_i}{n} > \epsilon\right) \leq \exp\left(-c\beta I(n\epsilon)\right) + n\exp\left(-I(n\epsilon)\right) = U_z(n,\epsilon)$
    $I(t) = c_1 \sqrt[\alpha]{t}, c_2 \log(t),\ c_2 > 2, \quad \lim_{n\to\infty} \frac{-\log \mathbb{P}\left(\sum z_i > n\epsilon\right)}{-\log U_z(n,\epsilon)} = 1$
  - Gaussian Region: $\frac{\epsilon'}{\sqrt{n}} < \epsilon_0$
    $\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum z_i > \epsilon'\right) \leq \exp\left(-\frac{\epsilon'^2}{2c_{n\epsilon_0,\beta}}\right) + n\exp\left(-\frac{n\epsilon_0^2}{\beta c_{n\epsilon_0,\beta}}\right)$
- $\epsilon = \epsilon(n)$ (see paper)

Sharpness

- Asymptotic Sharpness: Fix ϵ, $\quad n \to \infty$
  - First Region: $\epsilon \geq \epsilon_0$
    $\mathbb{P}\left(\frac{\sum z_i}{n} > \epsilon\right) \leq \exp\left(-c\beta I(n\epsilon)\right) + n\exp\left(-I(n\epsilon)\right) = U_z(n,\epsilon)$
    $I(t) = c_1 \sqrt[\alpha]{t}, c_2 \log(t),\ c_2 > 2, \qquad \lim_{n\to\infty} \frac{-\log \mathbb{P}\left(\sum z_i > n\epsilon\right)}{-\log U_z(n,\epsilon)} = 1$
  - Gaussian Region: $\frac{\epsilon'}{\sqrt{n}} < \epsilon_0$
    $\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum z_i > \epsilon'\right) \leq \exp\left(-\frac{\epsilon'^2}{2c_{n\epsilon_0,\beta}}\right) + n\exp\left(-\frac{n\epsilon_0^2}{\beta c_{n\epsilon_0,\beta}}\right)$
- $\epsilon = \epsilon(n)$ (see paper)

Sharpness

- Asymptotic Sharpness: Fix ϵ, $\quad n \to \infty$
  - First Region: $\epsilon \geq \epsilon_0$
    $\mathbb{P}\left(\frac{\sum z_i}{n} > \epsilon\right) \leq \exp\left(-c\beta I(n\epsilon)\right) + n\exp\left(-I(n\epsilon)\right) = U_z(n,\epsilon)$
    $I(t) = c_1\sqrt[\alpha]{t}, c_2\log(t),\ c_2 > 2, \quad \lim_{n\to\infty}\frac{-\log\mathbb{P}\left(\sum z_i > n\epsilon\right)}{-\log U_z(n,\epsilon)} = 1$
  - Gaussian Region: $\frac{\epsilon'}{\sqrt{n}} < \epsilon_0$
    $\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum z_i > \epsilon'\right) \leq \exp\left(-\frac{\epsilon'^2}{2c_{n\epsilon_0,\beta}}\right) + n\exp\left(-\frac{n\epsilon_0^2}{\beta c_{n\epsilon_0,\beta}}\right)$
- $\epsilon = \epsilon(n)$ (see paper)

Sharpness

- Asymptotic Sharpness: Fix $\epsilon, \quad n \to \infty$
  - First Region: $\epsilon \geq \epsilon_0$
    $\mathbb{P}\left(\frac{\sum z_i}{n} > \epsilon\right) \leq \exp\left(-c\beta I(n\epsilon)\right) + n\exp\left(-I(n\epsilon)\right) = U_z(n,\epsilon)$
    $I(t) = c_1 \sqrt[\alpha]{t}, c_2 \log(t), \; c_2 > 2, \qquad \lim_{n\to\infty} \frac{-\log \mathbb{P}\left(\sum z_i > n\epsilon\right)}{-\log U_z(n,\epsilon)} = 1$
  - Gaussian Region: $\frac{\epsilon'}{\sqrt{n}} < \epsilon_0$
    $\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum z_i > \epsilon'\right) \leq \exp\left(-\frac{\epsilon'^2}{2c_{n\epsilon_0,\beta}}\right) + n\exp\left(-\frac{n\epsilon_0^2}{\beta c_{n\epsilon_0,\beta}}\right)$
- $\epsilon = \epsilon(n)$ (see paper)

Proof Sketch

- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathbb{P}\left(\sum z_i^L > n\epsilon\right) + n\mathbb{P}\left(z > L\right)$
- Choose appropriate λ, L
  - $L = n\epsilon$ (Balance)
  - z_i^L subGaussian
  - $\mathrm{MGF}_{z^L}(\lambda) \xrightarrow{L\to\infty} \infty \quad (\forall \lambda > 0)$
  - Narrow curve $\lambda = \lambda(L)$

Proof Sketch

- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathbb{P}\left(\sum z_i^L > n\epsilon\right) + n\mathbb{P}\left(z > L\right)$
- Choose appropriate λ, L
  - $L = n\epsilon$ (Balance)
  - z_i^L subGaussian
    - $\left\|z_i^L\right\|_{\psi_2} \xrightarrow{L\to\infty} \infty$
  - $\mathrm{MGF}_{z^L}(\lambda) \xrightarrow{L\to\infty} \infty \quad (\forall\lambda > 0)$
  - Narrow curve $\lambda = \lambda(L)$
    - $\lambda = \frac{\beta I(L)}{L}$

Proof Sketch

- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathbb{P}\left(\sum z_i^L > n\epsilon\right) + n\mathbb{P}\left(z > L\right)$
- Choose appropriate λ, L
  - $L = n\epsilon$ (Balance)
  - z_i^L subGaussian
    - $\left\|z_i^L\right\|_{\psi_2} \xrightarrow{L\to\infty} \infty$
  - $\mathrm{MGF}_{z^L}(\lambda) \xrightarrow{L\to\infty} \infty \quad (\forall \lambda > 0)$
  - Narrow curve $\lambda = \lambda(L)$
    - $\lambda = \frac{\beta I(L)}{L}$

Proof Sketch

- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathbb{P}\left(\sum z_i^L > n\epsilon\right) + n\mathbb{P}\left(z > L\right)$
- Choose appropriate λ, L
  - $L = n\epsilon$ (Balance)
  - z_i^L subGaussian
    - $\left\|z_i^L\right\|_{\psi_2} \xrightarrow{L\to\infty} \infty$
  - $\mathrm{MGF}_{z^L}(\lambda) \xrightarrow{L\to\infty} \infty \quad (\forall\lambda > 0)$
  - Narrow curve $\lambda = \lambda(L)$
    - $\lambda = \frac{\beta I(L)}{L}$

Proof Sketch

- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathbb{P}\left(\sum z_i^L > n\epsilon\right) + n\mathbb{P}\left(z > L\right)$
- Choose appropriate λ, L
    - $L = n\epsilon$ (Balance)
    - z_i^L subGaussian
        - $\left\|z_i^L\right\|_{\psi_2} \xrightarrow{L\to\infty} \infty$
    - $\mathrm{MGF}_{z^L}(\lambda) \xrightarrow{L\to\infty} \infty \quad (\forall\lambda > 0)$
    - Narrow curve $\lambda = \lambda(L)$
        - $\lambda = \frac{\beta I(L)}{L}$

Proof Sketch

- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathbb{P}\left(\sum z_i^L > n\epsilon\right) + n\mathbb{P}\left(z > L\right)$
- Choose appropriate λ, L
  - $L = n\epsilon$ (Balance)
  - z_i^L subGaussian
    - $\left\| z_i^L \right\|_{\psi_2} \xrightarrow{L\to\infty} \infty$
  - $\mathrm{MGF}_{z^L}(\lambda) \xrightarrow{L\to\infty} \infty \quad (\forall \lambda > 0)$
  - Narrow curve $\lambda = \lambda(L)$
    - $\lambda = \frac{\beta I(L)}{L}$

Proof Sketch

- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n \epsilon}$
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathbb{P}\left(\sum z_i^L > n\epsilon\right) + n\mathbb{P}\left(z > L\right)$
- Choose appropriate λ, L
  - $L = n\epsilon$ (Balance)
  - z_i^L subGaussian
    - $\left\|z_i^L\right\|_{\psi_2} \xrightarrow{L\to\infty} \infty$
  - $\mathrm{MGF}_{z^L}(\lambda) \xrightarrow{L\to\infty} \infty \quad (\forall \lambda > 0)$
  - Narrow curve $\lambda = \lambda(L)$
    - $\lambda = \frac{\beta I(L)}{L}$

Proof Sketch

- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathrm{MGF}(\lambda)^n \mathrm{e}^{-\lambda n\epsilon}$
- $\mathbb{P}\left(\sum z_i > n\epsilon\right) \leq \mathbb{P}\left(\sum z_i^L > n\epsilon\right) + n\mathbb{P}\left(z > L\right)$
- Choose appropriate λ, L
  - $L = n\epsilon$ (Balance)
  - z_i^L subGaussian
    - $\left\|z_i^L\right\|_{\psi_2} \xrightarrow{L\to\infty} \infty$
  - $\mathrm{MGF}_{z^L}(\lambda) \xrightarrow{L\to\infty} \infty \quad (\forall\lambda > 0)$
  - Narrow curve $\lambda = \lambda(L)$
    - $\lambda = \frac{\beta I(L)}{L}$

Conclusion

- Concentration inequality for $\mathrm{Var}(z) < \infty$
- Ready to use (Only bound expectation of smooth function)
- Characterize Region of Gaussian deviation
- Asymptotically Sharp
- Many application in Machine learning and Data science

Conclusion

- Concentration inequality for $\mathrm{Var}(z) < \infty$
- Ready to use (Only bound expectation of smooth function)
- Characterize Region of Gaussian deviation
- Asymptotically Sharp
- Many application in Machine learning and Data science

Conclusion

- Concentration inequality for $\mathrm{Var}(z) < \infty$
- Ready to use (Only bound expectation of smooth function)
- Characterize Region of Gaussian deviation
- Asymptotically Sharp
- Many application in Machine learning and Data science

Conclusion

- Concentration inequality for $\mathrm{Var}(z) < \infty$
- Ready to use (Only bound expectation of smooth function)
- Characterize Region of Gaussian deviation
- Asymptotically Sharp
- Many application in Machine learning and Data science

Conclusion

- Concentration inequality for $\mathrm{Var}(z) < \infty$
- Ready to use (Only bound expectation of smooth function)
- Characterize Region of Gaussian deviation
- Asymptotically Sharp
- Many application in Machine learning and Data science